

Serial No. 09/837,158
Docket No. ARC920000018US1

2

AMENDMENTS TO THE CLAIMS:

1. (Previously presented) A computer-implemented method for identifying relationships between text documents and structured variables pertaining to said text documents, comprising:
 - generating a dictionary of keywords in said text documents;
 - forming categories of said text documents using said dictionary and an automated algorithm;
 - counting occurrences of said structured variables, said categories, and combinations of said structured variables and said categories for said text documents; and
 - calculating probabilities of occurrences of said combinations of structured variables and categories to identify a relationship between said text documents and said structured variables.
2. (Original) The method according to claim 1, wherein said algorithm comprises a keyword occurrence algorithm and wherein each of said categories comprises a category of text documents in which a particular keyword occurs.
3. (Original) The method according to claim 1, wherein said algorithm comprises a clustering algorithm and wherein each of said categories comprises a category of said text documents containing a particular cluster.
4. (Original) The method according to claim 3, wherein said clustering algorithm comprises a k means algorithm.
5. (Previously presented) The method according to claim 3, wherein said forming said categories comprises inputting a predetermined number of categories.
6. (Previously presented) The method according to claim 2, wherein said forming said categories comprises:
 - generating a sparse matrix array containing a count of each of said keywords in each of said text documents.

Serial No. 09/837,158
Docket No. ARC920000018US1

3

7. (Previously presented) The method according to claim 1, wherein said keywords comprise at least one of words and phrases which occur a predetermined number of times in said text documents.
8. (Original) The method of claim 1, wherein said calculating probabilities comprises using a Chi squared function.
9. (Original) The method of claim 1, wherein said generating a dictionary of keywords comprises:
 - first parsing text in said text document to identify and count occurrences of words;
 - storing a predetermined number of frequently occurring words;
 - second parsing text in said text documents to identify and count occurrences of phrases; and
 - storing a predetermined number of frequently occurring phrases.
10. (Original) The method according to claim 9, wherein said frequently occurring words and phrases are stored in a hash table.
11. (Original) The method according to claim 6, wherein said generating a sparse matrix array comprises:
 - third parsing text in said text documents to count a number of times that each of said keywords occurs in each of said text documents.
12. (Previously presented) The method according to claim 1, wherein said relationships comprise said combinations of structured variables and categories having a lowest probability of occurrence.
13. (Canceled)
14. (Previously presented) A computer-implemented method for identifying relationships between text documents and structured variables pertaining to said text documents,

Serial No. 09/837,158

4

Docket No. ARC920000018US1

comprising:

generating a dictionary of keywords in said text documents;

forming categories of said text documents using said dictionary and an automated algorithm;

counting occurrences of said structured variables, said categories, and combinations of said structured variables and said categories for said text documents; and

calculating probabilities of occurrences of said combinations of said structured variables and categories, said probabilities comprising a probability that a given co-occurrence of a structured variable and a category would have occurred as a purely random event.

15. (Original) The method according to claim 1, wherein said structured variables comprise predetermined time intervals.

16. (Original) The method according to claim 15, wherein said predetermined time intervals comprise one of days, weeks, months and years.

17. (Previously presented) A system for identifying relationships between text documents and structured variables pertaining to said text documents, comprising:

an input device for inputting text documents;

a processor for forming categories of said text documents and counting occurrences of said structured variables, categories, and combinations of said structured variables and said categories and calculating probabilities of

occurrence of said combinations of structured variables and categories to identify a relationship between said text documents and said structured variables ; and

a display for displaying said probabilities.

18. (Previously presented) The system according to claim 17, further comprising:

a memory for storing occurrences of said structured variables, categories, and combinations of structured variables and categories and probabilities of occurrences of said combinations of structured variables and categories .

Serial No. 09/837,158

5

Docket No. ARC920000018US1

19. (Original) The system according to claim 17, wherein said structured variables comprise predetermined time intervals.
20. (Original) The system according to claim 19, wherein said predetermined time intervals comprise one of days, weeks, months and years.
21. (Original) The system according to claim 17, wherein said system calculates a probability that a given co-occurrence of a structured variable and a category would have occurred as a purely random event.
22. (Original) The system according to claim 17, wherein said relationships comprise statistically significant relationships.
23. (Previously presented) A programmable storage medium tangibly embodying a program of machine-readable instructions executable by a digital processing apparatus to perform a method for identifying relationships between text documents and structured variables pertaining to said text documents, said method comprising:
- generating a dictionary of keywords in said text documents;
 - forming categories of said text documents using said dictionary and an automated algorithm;
 - counting occurrences of said structured variables, said categories, and combinations of said structured variables and said categories for in said text documents; and
 - calculating probabilities of occurrences of said combinations of structured variables and categories to identify a relationship between said text documents and said structured variables.
24. (New) The method according to claim 1, wherein said structured variables comprise structured data.